CLINICIAN ORIENTED ACCESS TO DATA - C.O.A.D.
A Natural Language Interface to a VA DHCP Database

Christine Levy, M.S.[1,3] and Elizabeth Rogers, M.D.[2]

[1]VA Quality Management Institute,[2]Department of Veterans Affairs GRECC, and [3]Duke - UNC Training Program in Medical Informatics, University of North Carolina in Chapel Hill

Abstract

Hospitals collect enormous amounts of data related to the on-going care of patients. Unfortunately, a clinicians access to the data is limited by complexities of the database structure and/or programming skills required to access the database. The COAD project attempts to bridge the gap between the clinical user's need for specific information from the database, and the wealth of data residing in the hospital information system. The project design includes a natural language interface to data contained in a VA DHCP database. We have developed a prototype which links natural language software to certain DHCP data elements, including, patient demographics, prescriptions, diagnoses, laboratory data, and provider information. English queries can by typed onto the system, and answers to the questions are returned. Future work includes refinement of natural language/DHCP connections to enable more sophisticated queries, and optimization of the system to reduce response time to user questions.

Introduction

VA information systems are written in M (MUMPS) which maintain data in a hierarchical data structure. The database at a typical VA hospital contains thousands of files, each with tens to hundreds of separate fields. Existing tools to access information are DHCP menu options, or VA Fileman. Menu options generate data reports which are specific to the ancillary package which they are part of, and are not normally customized for special information requests. Fileman is not easy to use, and requires an understanding of the DHCP data structure.

Natural language software allows English queries to be made of an underlying database. Software links are made between natural language and the database, so that relationships, entities, attributes and their interrelationships are known to the natural language processor. When a user types in a question, the natural language processor interprets the question, translates it into an SQL query, and then executes the query. The user may type in very general questions and does not need to know specifics about the database. If a question is not understood by the natural language processor, it guides them to assist in generating a question which can be answered.

Methods/Results

Natural Language Incorporated (NLI) software is used for the natural language interface. A VAX 4100 running VMS, NLI, Oracle and DECNET has been configured for the project. Portions of the DHCP database are extracted from the VA system and downloaded to the 4100. The data extraction routines are written in M, and run on the VA DHCP system every 15 minutes throughout the day. Data elements are extracted from the patient, prescription, laboratory, patient movement and provider files and loaded into Oracle tables. The VAX 4100 is connected to the hospital Ethernet backbone, therefore NLI can be accessed from any terminal in the hospital.

NLI software is linked to the underlying database by programming it to understand the relations in the database. It also has built in reasoning routines and knowledge of common English words and phrases. NLI programming tools were used to make the links between the database the English language.

We have successfully linked certain DHCP data elements to the natural language processor so that the user can type questions regarding the data. Simple questions are readily answered by the system. A simple question is one which asks information about an attribute of a relationship or joins information between two relationships. For instance, "Which patients are taking Digoxin?" is an example of a simple question. The system easily handles these questions regardless of how a user types the question. The same result is returned whether one asks: "Who is taking Digoxin?", "Which patients are being prescribed Digoxin?", "Digoxin is taken by whom?" and so on. Follow up questions are understood by the system. For example, "Who is treating them?" is interpreted by the system as: 'Of the patients taking Digoxin, who is their provider?'

Future work includes programming the system to respond to complex questions. More complex questions combine temporal aspects of data with parameters related to total care of the patient. For example, the series of questions 'Who has high glucose?', 'Were they started on insulin?', 'Did their glucose level decrease?' should be understood and appropriately answered by the system.

0195-4210/95/$5.00